GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 6, 2001, 19:32:01 ; Search time 182.97 Seconds
(without alignments)
7255.390 Million cell updates/sec

Title: US-09-494-297-1
Perfect score: 2274
Sequence: 1 atgaaaaaaacaaggtttcc.....ggataagaaaacatgactag 2274

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /cgn1_8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn1_8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn1_8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn1_8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn1_8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn1_8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn1_8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn1_8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn1_8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn1_8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn1_8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn1_8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn1_8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn1_8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn1_8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn1_8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn1_8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn1_8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn1_8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn1_8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn1_8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgn1_8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56.6	2.5	4677	21	A70259	Plasmodium falcipa
	2	56	2.5	11922	21	A70187	Plasmodium falcipa
	3	53.8	2.4	5940	21	A70105	Plasmodium falcipa
	4	51.4	2.3	1998	21	A70212	Plasmodium falcipa
	5	51.2	2.3	1527	21	A70121	Plasmodium falcipa
	6	51.2	2.3	3399	17	T05868	Chicken leucocytoz
c	7	50.2	2.2	15016	20	X99560	Nucleic acid seque
	8	49.6	2.2	3095	11	Q03875	Sequence encoding
	9	48.2	2.1	876	14	Q50947	Immunoglobulin bin
	10	48.2	2.1	3279	14	Q50946	Sequence encoding
	11	48.2	2.1	3279	14	Q51556	Sequence encoding
	12	48.2	2.1	20674	21	C58017	Arachidonic acid m
	13	47.4	2.1	5897	18	V74631	Staphylococcus aur
	14	47.4	2.1	7458	21	A70106	Plasmodium falcipa
	15	47.4	2.1	1664976	19	V21209	Methanococcus jann
	16	47	2.1	1062	20	X61556	B. burgdorferi ant
	17	47	2.1	1122	20	X61555	B. burgdorferi ant
c	18	47	2.1	1785	20	X99656	Nucleic acid seque
c	19	47	2.1	2418	20	X99561	Nucleic acid seque
	20	47	2.1	116277	20	X20249	Borrelia burgdorfe
	21	46.4	2.0	5361	18	T78868	P. falciparum live
	22	46.4	2.0	6152	18	T78867	P. falciparum live
	23	46.2	2.0	2503	15	Q53480	pNPX30 xylanase cD
c	24	46.2	2.0	910715	20	X20248	Borrelia burgdorfe
	25	46	2.0	3837	21	A70211	Plasmodium falcipa
	26	45.6	2.0	1686	16	Q87587	DNA encoding Leuco
	27	45.4	2.0	2841	18	V74488	Staphylococcus aur
	28	45.2	2.0	1938	17	T08079	Fibrinogen binding
c	29	44.8	2.0	997	21	A14996	cDNA encoding a hu
c	30	44.8	2.0	1220	18	V75229	Staphylococcus aur
c	31	44.8	2.0	3308	21	A26917	Essential Staphylo
	32	44.8	2.0	6042	21	A70199	Plasmodium falcipa
	33	44.6	2.0	3642	21	A70180	Plasmodium falcipa
	34	44.6	2.0	19124	18	T72882	Plasmodium var-7 g
	35	44.6	2.0	19124	21	Z98287	Plasmodium var-7 p
	36	44.4	2.0	3567	21	A70117	Plasmodium falcipa
c	37	44.4	2.0	5652	20	X99575	Nucleic acid seque
c	38	44.4	2.0	14066	20	X99556	Nucleic acid seque
c	39	44.4	2.0	26811	20	X20253	Borrelia burgdorfe
	40	44.2	1.9	6621	21	A70188	Plasmodium falcipa
c	41	44	1.9	1664976	19	V21209	Methanococcus jann
	42	43.8	1.9	5454	21	A70236	Plasmodium falcipa
	43	43.6	1.9	4311	21	A70133	Plasmodium falcipa
c	44	43.4	1.9	732	20	X99612	Nucleic acid seque
	45	43.4	1.9	1149	20	X99651	Nucleic acid seque

ALIGNMENTS

RESULT 1
A70259
ID A70259 standard; DNA; 4677 BP.
XX
AC A70259;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:392.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 565-566; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 4677 BP; 2106 A; 402 C; 966 G; 1203 T; 0 other;

Query Match 2.5%; Score 56.6; DB 21; Length 4677;
Best Local Similarity 43.5%; Pred. No. 0.0024;
Matches 416; Conservative 0; Mismatches 529; Indels 12; Gaps 3;

```
Qy   1084 tttaaggttgaagctggcaaagtgtatactattattgatggaaaacagattgaaaatccc 1143
Db   3256 tctgatttaaaagatcttgaagaagatatattaaaagaagtaaaagaaatcaaagaacctt 3315

Qy   1144 aataaagagatagtagagccttactcagtagaagcatataatgattttgaagaatttagc 1203
Db   3316 gaaagtgaaatttagaagattataaagaattaaaaactattgaaacagatattttagaa 3375

Qy   1204 gttttaactacacaaaactatgcaaaattttattatgcaaaaaataaaaatggaagttca 1263
Db   3376 gagaaaaaagaaatagaaaaagatcattttgaaaaattcgaagaagaagctgaagaaata 3435

Qy   1264 caggttgtctattgctttaatgcagatctaaaatctccaccagactctgaagatggtggg 1323
Db   3436 aaagatcttgaagcagatatattaaaagaagtatcttcattagaagttgaagaagaaaaa 3495

Qy   1324 aaaacaatgactccagactttacaacaggagaagtaaaatacactcatattgcaggtcgt 1383
Db   3496 aaattagaagaagtacacgaattaaaagaagaggtagaa------catataataagtggt 3549

Qy   1384 gacctctttaaatatactgtgaaaccaagagataccgatcctgacactttcttaaaacat 1443
Db   3550 gatgcgcatataaaaggtttggaagaagatgatttagaagaagtagatgatttaaaagga 3609

Qy   1444 atcaaaaaagtaattgagaagggttacagggaaaaaggacaagctattgagtatagtggt 1503
Db   3610 agtatattagacatgttaaagggagatatggaattaggggatatggataaggaaagttta 3669

Qy   1504 ctaactgagacacaattgcgtgcggctactcagttagcaatatattatttcactgatagt 1563
Db   3670 gaagatgtaacagcaaaacttggagaaagagttgaatccttaaaagatgttttatctagt 3729

Qy   1564 gctgaattagataaggataaactaaaagactatcatggttttggagacatgaatgatagt 1623
Db   3730 gc---attaggcatggatgaagaacaaatgaaaacaagaaaaaaagctcaaagacctaaa 3786

Qy   1624 actttagcagttgctaaaatccttgtagaatacgctcaagatagtaatcctccacagcta 1683
Db   3787 ttggaagaagtattattaaaagaagaggttaaagaagaaccaaagaaaaaaataacaaaa 3846

Qy   1684 actgaccttgatttctttattccgaataacaataaatatcaatctcttattggaactcag 1743
Db   3847 aagaaagtaaggtttgatattaaggataaggaaccaaaagatgaaatagtagaagttgaa 3906

Qy   1744 tggcatccagaagatttagttgatattattcgtatggaagataaaaaagaagttatacct 1803
Db   3907 atgaaagatgaagatatagatgaagatatagaagaagatgtagaagaagatatagaagaa 3966

Qy   1804 gtaactcataatttaacattgagaaaaacggtgactggtttagctggtgacagaactaaa 1863
Db   3967 gataaagttgaagatatagatgaagatatagatgaagatatagatgaagatataggtgaa 4026

Qy   1864 gatttccattttgaaattgaattaaaaaataataagcaagaattgctttctcaaactgtt 1923
Db   4027 gacaaagatgaagttatagatttaatagtccaaaaagagaaacgcattg---aaaaggtt 4083

Qy   1924 aaaacagataaaacaaacctcgaatttaaagatggtaaagcaaccattaatttaaaacat 1983
Db   4084 aaagagaaaaagaaaaaattagaaaaaaaagttgaagaaggtgttagtggtcttaaaaaa 4143

Qy   1984 ggggaaagtttaacacttcaaggtttaccagaaggttattcttaccttgtcaaagaa 2040
Db   4144 cacgtagacgaagtaatgaaatatgttcaaaaaattgataaagaagttgataaagaa 4200
```

RESULT 2
A70187
ID A70187 standard; DNA; 11922 BP.
XX
AC A70187;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:320.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 516-519; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum

```
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 11922 BP; 5402 A; 948 C; 1343 G; 4229 T; 0 other;


  Query Match             2.5%;  Score 56;  DB 21;  Length 11922;
  Best Local Similarity   41.3%;  Pred. No. 0.0041;
  Matches  380;  Conservative    0;  Mismatches  540;  Indels    0;  Gaps    0;

Qy     988 aatgatattggagaaagaattgaactatcagatggaacttatactttaactgaattgaat 1047
           |||   | |  | | | || |       |||||     | ||  |  |  | |    |||
Db   10261 aatacaaatacaaataaaaattctgatacagataatcataatgatgcatataatcataat 10320

Qy    1048 tctccagctggttatagtatcgcagagccaatcacttttaaggttgaagctggcaaagtg 1107
             | ||  |  | ||| |   ||| |       | |        | |   |    | |
Db   10321 gatgcatataatcataatgatgcatataatcataatgacgcatataatcataatgatgca 10380

Qy    1108 tatactattattgatggaaaacagattgaaaatcccaataaagagatagtagagccttac 1167
           |||| |  || |||||  | |  |   | |  || |  ||||  | |  |  |    | |
Db   10381 tataatcataatgatgcatataatcataatgatgcatataatcataatgatgcatataat 10440

Qy    1168 tcagtagaagcatataatgattttgaagaatttagcgttttaactacacaaaactatgca 1227
             || |||||||||| ||  ||| | |  ||     |     |||| | ||  ||
Db   10441 cataatgatgcatataatcataatgatgcagataatcataatgatacagataatcatagt 10500

Qy    1228 aaattttattatgcaaaaaataaaaatggaagttcacaggttgtctattgctttaatgca 1287
             |   |||| || |  | ||  | || |||| ||   |  ||     |     | |||
Db   10501 gataattataattctcataaatataaaggaacttataaaatttatcgtattcatgatgag 10560

Qy    1288 gatctaaaatctccaccagactctgaagatggtgggaaaacaatgactccagactttaca 1347
           ||    |  || |    |  | |      |    |     | | || | | ||
Db   10561 gaagatattattcaagataataattatacaaacgatgattatgttaatacaaattataat 10620

Qy    1348 acaggagaagtaaaatacactcatattgcaggtcgtgacctctttaaatatactgtgaaa 1407
             ||  ||  |  ||| |           ||       |  |  |   |
Db   10621 tttcatcaaaataataattatcaaaacagctccaaggataatagccatacttatactagt 10680

Qy    1408 ccaagagataccgatcctgacactttcttaaaacatatcaaaaaagtaattgagaagggt 1467
           |  | |||    ||    | | |  || | |  | | || |||| || |
Db   10681 tcgggtgatgataatttaaataatccatttatatctttaaataaagaaaatatatccaaa 10740

Qy    1468 tacagggaaaaaggacaagctattgagtatagtggtctaactgagacacaattgcgtgcg 1527
           | |  ||||  | || | | | |  |||| |    |   | | | |      ||
Db   10741 aaaaaaaaaaacaaaaaaacaaataaaaatagagaagaagaaaaaaaaaagaaaaatgaa 10800

Qy    1528 gctactcagttagcaatatattatttcactgatagtgctgaattagataaggataaacta 1587
             ||  ||    | | || | ||||||   ||   |  |  |   | |   |  ||| ||
Db   10801 attatacataatgaagtagaatatttcttagaagctattttaaatgttgtagggaaatta 10860

Qy    1588 aaagactatcatggttttggagacatgaatgatagtactttagcagttgctaaaatcctt 1647
           |||   |  |  ||||  |    | ||       |   |      |      ||  ||
Db   10861 tcagatcgtattaattttatatttttaaaaatcgaaaaaattaattgtctagctatttttt 10920

Qy    1648 gtagaatacgctcaagatagtaatcctccacagctaactgaccttgatttctttattccg 1707
           | ||||| |  || | ||  | | |  | |     ||       | |||    ||||
Db   10921 gaagaatcctatctatatgttgaactactaacaaaaatatcttattattgtaatatttat 10980

Qy    1708 aataacaataaatatcaatctcttattggaactcagtggcatccagaagatttagttgat 1767
           ||||  ||||| |    |     | |      | |  |   |  | |  ||||| || ||
Db   10981 aatataaataatttcttaatatctttaatggataaaagagttgaacatcatttattttat 11040

Qy    1768 attattcgtatggaagataaaaaagaagttatacctgtaactcataatttaacattgaga 1827
           || | | | ||   ||| |||           |   | |  |  ||  |   |||  ||

Db   11041 ataaatagaatattagaaaaattcatttacgaaaaagaagttattatatatccatatagt 11100

Qy    1828 aaaacggtgactggtttagctggtgacagaactaaagatttccattttgaaattgaatta 1887
           |      | | |  |  | |     |   | ||||  ||    ||    |||   ||  |
Db   11101 ccttatgaaattaatgcatcaacaaatgtacctaataatcaagataaaaaaaaaaaaaaa 11160

Qy    1888 aaaaatаataagcaagaatt 1907
           |||   |  ||  | ||| |
Db   11161 aaatcaacaaaaaatgaact 11180


RESULT   3
A70105
ID   A70105 standard; DNA; 5940 BP.
XX
AC   A70105;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:238.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 460-462; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 5940 BP; 3106 A; 343 C; 879 G; 1612 T; 0 other;
```

```
  Query Match             2.4%;  Score 53.8;  DB 21;  Length 5940;
  Best Local Similarity   44.7%;  Pred. No. 0.011;
  Matches  294;  Conservative    0;  Mismatches  357;  Indels    6;  Gaps    2;

Qy   1477 aaaggacaagctattgagtatagtggtctaactgagacacaattgcgtgcggctactcag 1536
          || |   ||| || |||  ||| ||     |  || | |           |  ||   |
Db     73 aatgaggaagataatgaaaataatgaaaagagggaaaaaagcgattcatggtataaaaaa 132

Qy   1537 ttagcaatatattatttcactgatagtgctgaattagataaggataaactaaaagactat 1596
           ||  |  |  | |    |    ||   || |||   | || ||||      |||  ||
Db    133 ataatagaaactaaaggaaaaagtaaaactaaatataaaaatgataattccttagatgat 192

Qy   1597 catggttttggagacat---gaatgatagtactttagcagttgctaaaatccttgtagaa 1653
           ||  |  ||  |||||    ||| ||| || |       |  | | |    |   ||
Db    193 aatattaatgaggacataataaataataataataataataatgataataataatgat 252

Qy   1654 tacgctcaagatagtaatcctccacagctaactgaccttgatttctttattccgaataac 1713
           ||  | | |||| ||||  |    |    | |||   | ||     || |   ||| |
Db    253 aacaataatgataataataatgataataataatgataataataatgataataataatgag 312

Qy   1714 aataaatatcaatctcttattggaactcagtggcatccagaagatttagttgatattatt 1773
          |||||  || |   |  || |   | | |         |||   | |   |||||||
Db    313 aataataatgataataataatttttaataattatagcgatgaaatatcaaaaaatattata 372

Qy   1774 cgtatggaagataaaaaagaagttatacctgtaactcataatttaacattgagaaaaacg 1833
           | ||  ||  || |   ||||      | |     |          |  | ||      |
Db    373 cataaagacaatgagctagaaaaccagcttaaggatacattaaagtccattagttcgttg 432

Qy   1834 gtgactggtttagctggtgacagaactaaagatttccattttgaaattgaattaaaaaat 1893
            || |    | |    | ||  || ||||  |    | || ||||  ||||||||| |
Db    433 tcgaataaaattgtgaattacgaaagtaaaattgaagaattagaaaaagaattaaaagaa 492

Qy   1894 aataagcaagaattgctttctcaaactgttaaaacagataaaacaaacctcgaatttaaa 1953
            ||| |  |     ||  | | | || | |   | |||||  |||     ||    |
Db    493 gtaaaggataagaatattgataataatgattatgaaaataaattaaaagaaaaagaagat 552

Qy   1954 gatggtaaagcaaccattaatttaaaacatggggaaagtttaacacttcaaggtttacca 2013
            || ||||  ||  ||| || |   | ||   ||||    ||  ||| |     |  |
Db    553 tttgttaaacaaaaaattgatatgctaaat---gaaaaagaaaatcttttacaagaaaaa 609

Qy   2014 gaaggttattcttaccttgtcaaagaaacagattctgaaggctataaggttaaagttaat 2073
          |||    ||  | |       || | | |  |  ||||    | |||  || | | ||
Db    610 gaattagatattaataaaagagaaaagaaaattaatgaaaaagaaaagaatatataataaa 669

Qy   2074 agccaagaagtagcaaatgctacagtttcaaaaacaggaataacaagtgatgagaca 2130
          |   |||||  |    ||  || ||    | |       |  || || | | || | |
Db    670 aaggaagaaacatttcataatatagaaaaagagtatttagaaaaaaataaagaaaga 726


RESULT   4
A70212
ID   A70212 standard; DNA; 1998 BP.
XX
AC   A70212;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:345.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 537; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 1998 BP; 1127 A; 141 C; 338 G; 392 T; 0 other;


  Query Match             2.3%;  Score 51.4;  DB 21;  Length 1998;
  Best Local Similarity   41.8%;  Pred. No. 0.027;
  Matches  453;  Conservative    0;  Mismatches  626;  Indels    6;  Gaps    2;

Qy   1115 ttattgatggaaaacagattgaaaatcccaataaagagatagtagagccttactcagtag 1174
          ||| | |||||| ||| | |     |    | ||    |   | |   ||   || |
Db    239 ttacttatggaacacatactagctttttacaaaactgcaccatcaatgattgtgtagatg 298

Qy   1175 aagcatataatgattttgaagaatttagcgttttaactacacaaaactatgcaaaatttt 1234
             |   |||| |||| ||||     ||   |   ||    | || |  ||   ||   |
Db    299 tcgataataaagattctgaaattaataatatatacaaaagaaaaagatgataataataata 358

Qy   1235 attatgcaaaaaataaaaatggaagttcacaggttgtctattgctttaatgcagatctaa 1294
          || ||| ||  ||  ||| || |     | |     || ||       ||   ||  |
Db    359 ataatggaactaagcaaattgaagagaaaaataaaatcaataaaagcgatttgcatagac 418

Qy   1295 aatctccaccagactctgaagatggtgggaaaacaatgactccagactttacaacaggag 1354
          ||  |  | |  |  | |  |||| ||||                 || ||  | |
Db    419 aaaatgaattaaa---tttacaaagtggaaaaaatgaacaagacataaataaaaatgaaa 475

Qy   1355 aagtaaaatacactcatattgcaggtcgtgacctctttaaatatactgtgaaaccaagag 1414
          ||| |||| |   |     |        ||        ||| |   ||| |||  | |||
Db    476 aaggaaaacaggatatctctaattccaatgcagagaacaaaaaggatgtaaaagaaggag 535

Qy   1415 ataccgatcctgacactttcttaaaacatatcaaaaaagtaattgagaagggttacaggg 1474
            |  ||    ||         ||||  |    ||||    |  |||  |         ||
Db    536 tgaaagaattagaagaaaagaaaaaagaagagaaaatttcagatgatcataaagtagagg 595

Qy   1475 aaaaaggacaagctattgagtatagtggtctaactgagacacaattgcgtgcggctactc 1534
          ||||   | || ||  |||  |||  | |       | | | |||    ||   ||
Db    596 aaaacaaaaaatctgatgaccataaagttgaggaaaacaaaaaatctgatgatcataaag 655

Qy   1535 agttagcaatatattatttcactgatagtgctgaattagataaggataaactaaaagact 1594
```

```
Db    656 ttgaggaaaacaaaaaatctgatgatcataaaatagaagaagtaaaaaaagtagaggaac 715

Qy   1595 atcatggttttggagacatgaatgatagtactttagcagttgctaaaatccttgtagaat 1654

Db    716 atgaagaagacgaagaagaggataagaaagaaaaaaaatcagaaaacaaaaataaagatg 775

Qy   1655 acgctcaagatagtaatcctccacagctaactgacct---tgatttctttattccgaata 1711

Db    776 aaaataaagatgaaaatgatgaagataatgatgaaataagtgatgaagatgaagttgatg 835

Qy   1712 acaataaatatcaatctcttattggaactcagtggcatccagaagatttagttgatatta 1771

Db    836 atgatgttgaagaagataaaaatgaaaatgatgatattgatgatgataaaaaggaaacag 895

Qy   1772 ttcgtatggaagataaaaaagaagttatacctgtaactcataatttaacattgagaaaaa 1831

Db    896 ataaaacacatttggaagaagaggaaaatgaaataattgaaaaggaatttagtgataaga 955

Qy   1832 cggtgactggtttagctggtgacagaactaaagatttccattttgaaattgaattaaaaa 1891

Db    956 aaaagaatggtaaaaataaagatactaagaaggaaaaaagtaaagatactgagaaggaaa 1015

Qy   1892 ataataagcaagaattgctttctcaaactgttaaaacagataaaacaaacctcgaattta 1951

Db   1016 aaagtaaagatatagagaaggaaaaaagtaaagataaagagaaggaaaaaagtaaagata 1075

Qy   1952 aagatggtaaagcaaccattaatttaaaacatggggaaagtttaacacttcaaggtttac 2011

Db   1076 aagagaaggaaaaaggtaaagataaagagaaggaaaaaagtaaagatattgagaaggaaa 1135

Qy   2012 cagaaggttattcttaccttgtcaaagaaacagattctgaaggctataaggttaaagtta 2071

Db   1136 aagaaaaagataaagatatagagaaggaaaaaagtaaagatactgcgaaggaaaaagaaa 1195

Qy   2072 atagccaagaagtagcaaatgctacagtttcaaaaacaggaataacaagtgatgagacac 2131

Db   1196 aagataaagatatagagaaagaaaaaagcaaagatatggaaaaattaaagaataaacaaa 1255

Qy   2132 ttgcttttgaaaataataaagagcctgttgttcctacaggagttgatcaaaagatcaatg 2191

Db   1256 atgatgaaaaaaagaaagacgataacgaaaagaaaaaaaatgataaacaagatatacatg 1315

Qy   2192 gctat 2196

Db   1316 atgat 1320


RESULT   5
A70121
ID   A70121 standard; DNA; 1527 BP.
XX
AC   A70121;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:254.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 472; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 1527 BP; 848 A; 94 C; 210 G; 375 T; 0 other;


  Query Match             2.3%;  Score 51.2;  DB 21;  Length 1527;
  Best Local Similarity   45.2%;  Pred. No. 0.028;
  Matches  228;  Conservative    0;  Mismatches  273;  Indels    3;  Gaps    1;

Qy   1698 ctttattccgaataacaataaatatcaatctcttattggaactcagtggcatccagaaga 1757

Db    294 ctgtactcctaataataataaaatatttttttataatgtccatttaataaatgatgatga 353

Qy   1758 tttagttgatattattcgtatggaagataaaaaagaagttatacctgtaactcataattt 1817

Db    354 tctttttaagagaagaaaaaataaaaaaaagaaaaagaaaatgataaccttaaagataaa 413

Qy   1818 aacattgagaaaaacggtgactggtttagctggtgacagaactaaagatttccattttga 1877

Db    414 taaatgtaaccataaagataaaaatttacataaaaatgagatgaaagatggtgatcatg- 472

Qy   1878 aattgaattaaaaaatataagcaagaattgctttctcaaactgttaaaacagataaaac 1937

Db    473 --ttttttcatatactaaaaagaagtggttgaataataataataataataataatattac 530

Qy   1938 aaacctcgaatttaaagatggtaaagcaaccattaatttaaaacatggggaaagtttaac 1997

Db    531 aaatatggtctctttttttgggttatggaaatataaaaagaaagtatgtaacaaataaatg 590

Qy   1998 acttcaaggtttaccagaaggttattcttaccttgtcaaagaaacagattctgaaggcta 2057

Db    591 tattataaatgaacaagaaaataataagatggatgagaatcaacatattgataaaaataa 650

Qy   2058 taaggttaaagttaatagccaagaagtagcaaatgctacagtttcaaaaacaggaataac 2117

Db    651 aaatattaatattaatatcaatttacatgatgataagaacgatgaaataagaaaacattc 710

Qy   2118 aagtgatgagacacttgcttttgaaaataataaagagcctgttgttcctacaggagttga 2177

Db    711 aaccatacaaacactttatcatagtaataataaagaaaagataatttcaaaaaatgttct 770
```

```
Qy   2178 tcaaaagatcaatggctatctagc 2201
          || |    | |    || || |
Db    771 aaaagatgaaagtacaaatataac 794


RESULT   6
T05868
ID   T05868 standard; DNA; 3399 BP.
XX
AC   T05868;
XX
DT   14-AUG-1996  (first entry)
XX
DE   Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
KW   Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
KW   immunisation; vaccination; ss.
XX
OS   Chicken leucocytozoan.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3399
FT                   /*tag=  a
FT   misc_feature    1150..3218
FT                   /*tag=  b
FT                   /note= "fragment referred to in the claims, for
FT                          use as insert in a recombinant vaccine
FT                          against chicken leucocytozoan disease"
XX
PN   JP07284392-A.
XX
PD   31-OCT-1995.
XX
PF   19-APR-1994;   94JP-0080643.
XX
PR   19-APR-1994;   94JP-0080643.
XX
PA   (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA   (KITA ) KITASATO KENKYUSHO SH.
XX
DR   WPI; 1996-006311/01.
DR   P-PSDB; R97866.
XX
PT   Chicken leucocytozoan immunogenic protein  - used in a recombinant
PT   vaccine against chicken leucocytozoan disease
XX
PS   Claim 6; Page 6-9; 35pp; Japanese.
XX
CC   T05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
CC   or a fragment of it can be used in a recombinant vaccine to immunise
CC   against chicken leucocytozoan disease. The DNA is used in a vector
CC   and operatively linked to an expression regulatory sequence as in
CC   standard practice.
XX
SQ   Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;


  Query Match             2.3%;  Score 51.2;  DB 17;  Length 3399;
  Best Local Similarity   41.9%;  Pred. No. 0.034;
  Matches  445;  Conservative    0;  Mismatches  608;  Indels    9;  Gaps    2;

Qy   1092 tgaagctggcaaagtgtatactattattgatggaaaacagattgaaaatcccaataaaga 1151
          ||||   |  | | | ||||  |  ||||  |  | |  |  |       |  |||
Db   1542 tgaaagagaaataatttatatagtagatgatttaccagaagtaaatgtagaagaatcaga 1601

Qy   1152 gatagtagagccttactcagtagaagcatataatgattttgaagaatttagcgttttaac 1211
            | ||| | |    ||   |||  | |||||||  |  |||||      |    |
Db   1602 tgaaacagaacatataacatatgaaatagataatgacatacaagaagagcatgaaaaagt 1661

Qy   1212 tacacaaaactatgcaaaattttattatgcaaaaaataaaaatggaagttcacaggttgt 1271
          |||||  |  | | ||||    |     || |  |  |||  | ||    | ||  ||
Db   1662 aacacatgaagaagaaaaagaagaagtaacacatgaagaaatagaaaaagaagagcatga 1721

Qy   1272 ctattgctttaatgcagatctaaaatctccaccagactctgaagatggtgggaaaacaat 1331
            |     | ||| |||   ||||     |  |     ||||||    |  |||  |
Db   1722 agaagtaatacatgaagaagaaaaagaagaagtaacacatgaagaaatagaaaaagaaga 1781

Qy   1332 gactccagactttacaacaggagaagtaaaatacactcatattgcaggtcgtgacctctt 1391
          |  |  |||  | | |   | ||||| |||| |       |     |       |
Db   1782 gcatgaagaagtaatacatgaagaagaaaaagaagaagtaacacatgaagaaaaagaaaa 1841

Qy   1392 taaatatactgtgaaaccaagagataccgatcctgacactttcttaaaacatatcaaaaa 1451
            || |   ||   ||  || | ||   ||     |        ||| ||||    |||
Db   1842 agaagagcatgaagaagtaatacatgaagaagaaaaagaagaagtaacacatgaagaaat 1901

Qy   1452 agtaattgagaagggttacagggaaaaaggacaagctattgagtatagtggtctaactga 1511
          || ||  ||  ||  | |    | || |    |||      |  |    |   |  |||
Db   1902 agaaaaagaagagcatgaagaagtaatacatgaagaagaaaaagaagaagtaacacatga 1961

Qy   1512 gacacaattgcgtgcggctactcagttagcaatatattatttcactgatagtgctgaatt 1571
            |  |      |  |    | |   || |||| || |       |    |  | |
Db   1962 agaaatagaaaaagaagagcatgaagaagtaatacatgaggaagaaaaagaagaagtaac 2021

Qy   1572 agataaggataaactaaaagactatcatggttttggagacatgaatgatagtactttagc 1631
          | || | || | |  ||||||  | ||||     | |       | ||    |   ||
Db   2022 acatgaagaaatagaaaaagaagagcatgaagaagtaatacatgaagaagaaaaagaaga 2081

Qy   1632 agttgctaaaatccttgtagaatacgctcaagatagtaatcctccacagctaactgacct 1691
          |||  |  |       |||||  | |   |  ||    |      |||   |   ||
Db   2082 agtaacacatgaagaaatagaaaaagaagagcatgaagaagtaatacatgaagaagaaaa 2141

Qy   1692 tgatttctttattccgaataacaataaatatcaatctcttattggaactcagtggcatcc 1751
            ||     | |  |   |  | ||  || |  ||   | |   | |         |||
Db   2142 agaagaagtaacacatgaagaaaaagaaaaagaagagcatgaagaagt---aatacatga 2198

Qy   1752 agaagatttagttgatattattcgtatggaagataaaaaagaagttatacctgtaactca 1811
          ||||||   ||  ||  | |  | |   |||||  ||  || |||| | || || |    |
Db   2199 agaagaaaaagaagaagtaacacatgaagaagaagaagaaaaagtaacacatgaagaaga 2258

Qy   1812 taatttaacattgagaaaaacggtgactggtttagctggtgacagaactaaagatttcca 1871
            |   || | | | | |    |          |   |  ||  ||   | ||     |
Db   2259 agaagaaaaagtaatacatgaagaaga------aaaagaagaagtaatacatgaagaaga 2312

Qy   1872 ttttgaaattgaattaaaaaataataagcaagaattgctttctcaaactgttaaaacaga 1931
            |||   | |  | |  |  | |  |||||    |  | ||    |  ||| |||
Db   2313 aaaagaagaagtaacacatgaagaagaaaaagaagaagtaacacatgaagaagaaaaaga 2372

Qy   1932 taaaacaaacctcgaatttaaagatggtaaagcaaccattaatttaaaacatggggaaag 1991
            ||  ||  |  |||    || | |   ||| |||   | |   | || | |  |||
Db   2373 agaagtaacacatgaagaagaaaaagaagaagtaacacatgaagaagaaaaagaagaagt 2432

Qy   1992 tttaacacttcaaggtttaccagaaggttattcttaccttgtcaaagaaacagattctga 2051
            |       |||    |  || |    ||    |        |||||  |     |||
Db   2433 aacacatgaagaagaagaaaaagtaacacatgaagaagaaaaagaagaagtaacacatga 2492

Qy   2052 aggctataaggttaaagttaatagccaagaagtagcaaatgctacagtttcaaaaacagg 2111
          ||   | || |   |||| |      ||||||  || ||| |  |||  |  | ||  ||
Db   2493 agaagaaaaagaagaagtaacacatgaagaagaagaaaaagtaacacatgaagaagaaga 2552

Qy   2112 aataacaagtgatgagacacttgctttttgaaaataataaaga 2153
          || |  ||   ||||   |   |   | ||   || ||||
Db   2553 aaaagtaacacatgaagaagaagaaaaagtaacatatgaaga 2594


RESULT   7
X99560/c
ID   X99560 standard; DNA; 15016 BP.
XX
AC   X99560;
XX
DT   05-OCT-1999  (first entry)
XX
DE   Nucleic acid sequence from U. urealyticum.
XX
KW   Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
```

KW human urogenital tract; pregnancy; neonatal disease; drug therapy;
KW suppurative arthritis; ss.
XX
OS Ureaplasma urealyticum.
XX
PN WO9939007-A1.
XX
PD 05-AUG-1999.
XX
PF 29-JAN-1999; 99WO-US01972.
XX
PR 30-JAN-1998; 98US-0073189.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Cassell GH, Chen EY, Glass JI, Glass JS, Heiner CR;
PI Lefkowitz E;
XX
DR WPI; 1999-469343/39.
XX
PT Detection of Ureaplasma urealyticum using novel genes, probes and
PT primers
XX
PS Claim 1; Page 48-53; 110pp; English.
XX
CC The present invention provides methods for the detection and diagnosis
CC of Ureaplasma urealyticum infection. It provides novel genes (X99501-681)
CC that can be used as a source of primers and probes for the detection and/
CC or quantification of U. urealyticum in a biological sample. The probes
CC that can be used in the method of the invention by forming target:probe
CC complex is copmplementary to a region selected from one of the 181
CC nucleic acid sequences (X99501-681). U. urealyticum is an opportunistic
CC pathogen of the human urogenital tract that is a significant cause of
CC adverse pregnancy outcome, neonatal disease, and suppurative arthritis.
CC As the infections are commonly asymptomatic, it is important to have
CC specific and sensitive methods for detecting their presence in a patient.
CC Also, as the pathogen has no current antibiotic directed specifically
CC against it, it would be advantageous to isolate and detect gene sequences
CC which are unique to it, and utilise these as a basis for diagnosis of
CC U. urealyticum infection as well as to develop new and improved drug
CC therapies. The present invention provides such novel polynucleotide
CC sequences (X99501-681).
XX
SQ Sequence 15016 BP; 4931 A; 1914 C; 1794 G; 6377 T; 0 other;

Query Match 2.2%; Score 50.2; DB 20; Length 15016;
Best Local Similarity 47.5%; Pred. No. 0.084;
Matches 182; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

Qy 1781 aagataaaaaagaagttatacctgtaactcataatttaacattgagaaaaacggtgactg 1840
Db 7274 ATGTTAAATATGCTTTATTTAATCAAACGCAAAATGCTAATTTAATTAACAATATTACAA 7215

Qy 1841 gtttagctggtgacagaactaaagatttccattttgaaattgaattaaaaaataataagc 1900
Db 7214 TTAAATATGATCAACAAAATAAACAATTCTATTGTGAATTTGATATTGATAATTTAACTT 7155

Qy 1901 aagaattgctttctcaaactgttaaaacagataaaacaaacctcgaatttaaagatggta 1960
Db 7154 TAAATCAAGATTATGGTATTTTTGAAATTAGTTTTGCTAATAAACCAATTCATGCCGCAT 7095

Qy 1961 aagcaaccattaatttaaaacatggggaaagtttaacacttcaaggtttaccagaaggtt 2020
Db 7094 TTGCTAAAATTAATGATACCCA---ACATCCTTTAGCTCTTTTAAGTACTAATAATCAAG 7038

Qy 2021 attcttaccttgtcaaagaaacagattctgaaggctataaggttaaagttaatagccaag 2080
Db 7037 ATATAAATCAAAAAATTGCTTTAGATTGTGAAGTTAATAAAGTTACAATTAATAAACAAG 6978

Qy 2081 aagtagcaaatgctacagtttcaaaaacaggaataacaagtgatgagacacttgcttttg 2140
Db 6977 ATCTTGATTATGATCAAAACACCAAGATTAGTTGACCGATTGAATTAAGTTCTAATTTAT 6918

Qy 2141 aaaataataaagagcctgttgtt 2163
Db 6917 TAAACAATAAGTACATTCGTGTT 6895

RESULT 8
Q03875
ID Q03875 standard; DNA; 3095 BP.
XX
AC Q03875;
XX
DT 24-AUG-1990 (first entry)
XX
DE Sequence encoding carboxylic terminal part of native GLURP.
XX
KW Plasmodium falciparum; antigen; malaria; vaccine; GLURP;ss
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..2352
FT /*tag= a
FT /product=GLURP
XX
PN WO9022811-A.
XX
PD 22-MAR-1990.
XX
PF 18-SEP-1989; 89WO-0000218.
XX
PR 03-MAR-1989; 89US-0218885.
PR 03-MAR-1989; 89DK-0005191.
XX
PA (STAT-) STATENS SERUMINST.
XX
PI Dziegiel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A;
PI Jakobsen PH;
XX
DR WPI; 1990-115998/15.
DR P-PSDB; R05804.
XX
PT Polypeptide(s) derived from Plasmidium falciparum antigen - used in
PT vaccines and in production of antibodies, for diagnosis and
PT therapy of malaria.
XX
PS Disclosure; Fig 7; 108pp; English.
XX
CC An open reading frame of 2349 bps extends from the 5' terminal end of the
CC insert to a "TAA" stop codon. It is longest ORF found in the sequence.
CC Sequence displays some of the characteristics of other malaria nucleic
CC acid sequences : tandemly repeated motifs, high AT content and a
CC corresponding preference for codons containing these bases, and a high
CC content of codons for glutamate. Three major repetitive sequences are:
CC one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
CC motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp
CC 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is
CC repeated tandemly 11 times.This last repetitious region consists of 360bp
CC repeats differing only in 3 bases GAT coding for aspartate. This region
CC is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content
CC of the coding part of the insert is on average 30%, and of the non-coding
CC 3' terminal 11% .
XX
SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T; 0 other;

Query Match 2.2%; Score 49.6; DB 11; Length 3095;
Best Local Similarity 42.9%; Pred. No. 0.075;
Matches 320; Conservative 0; Mismatches 414; Indels 12; Gaps 1;

Qy 1453 gtaattgagaagggttacagggaaaaaggacaagctattgagtatagtggtctaactgag 1512
Db 1057 gaaattgaggaagtttttcctgaaccaaatcaaaataacgaatttcaagaaattaatgaa 1116

```
Qy   1513 acacaattgcgtgcggctactcagttagcaatatattatttcactgatagtgctgaatta 1572
Db   1117 gatgataaaagtgcacatattcagcatgaaatagtagaagtagaagaaatacttccagaa 1176

Qy   1573 gataaggataaactaaaagactatcatggttttggagacatgaatgatagtactttagca 1632
Db   1177 gatgataaaaatgaaaaagttgaacatgaaatagtagaagttgaagaaattctagcagaa 1236

Qy   1633 gttgctaaaatccttgtagaatacgctcaagatagtaatcctccacagctaactgacctt 1692
Db   1237 gataaaaatgaaaaaggtcaacatgaaatagtagaggttgaagaaattctaccagaagat 1296

Qy   1693 gatttctttattccgaataacaataaatatcaatctcttattggaactcagtggcatcca 1752
Db   1297 gataaaaatgaaaaaggtcaacatgaaatagtagaagttgaagaaattctaccagaagat 1356

Qy   1753 gaagatttagttgatattattcgtatggaagataaaaaagaagttatacctgtaactcat 1812
Db   1357 aaaaatgaaaaaggtcaacatgaaatagtagaggttgaagaaattctaccagaagataaa 1416

Qy   1813 aatttaacattgagaaaaacggtgactggtttagctggtgacagaactaaagatttccat 1872
Db   1417 aatgaaaaagttgaacatgaaatagtagaagttgaagaaattctaccagaagataaaaat 1476

Qy   1873 tttgaaattgaattaaaaaatataagcaagaattgctttctcaaactgttaaaacagat 1932
Db   1477 gaaaaaggtcaacatgaaatagtagaggttgaagaaattctaccagaagataaaaatgaa 1536

Qy   1933 aaaacaaacctcgaatttaaagatggtaaagcaaccattaatttaaaacatggggaaagt 1992
Db   1537 aaagttcaacatgaaatagtagaagttgaagaaattctaccagaagataaaaatgaaaaa 1596

Qy   1993 ttaacacttcaaggtttaccagaaggttattcttaccttgtcaaagaaacagattctgaa 2052
Db   1597 ggtcaacatgaaatagtagaggttgaagaaattctaccagaagaagataaaaatgaaaaa 1656

Qy   2053 ggcta----------taaggttaaagttaatagccaagaagtagcaaatgctacagtt 2100
Db   1657 ggtcaacatgaaatagtagaggttgaagaaattctaccagaagataaaaatgaaaaagtt 1716

Qy   2101 tcaaaaacaggaataacaagtgatgagacacttgcttttgaaaataataaagagcctgtt 2160
Db   1717 caacatgaaatagtagaggttgaagaaattctaccagaagataaaaatgaaaaagttcaa 1776

Qy   2161 gttcctacaggagttgatcaaaagat 2186
Db   1777 catgaaatagtagaggttgaagaaat 1802
```

RESULT 9
Q50947
ID Q50947 standard; cDNA; 876 BP.
XX
AC Q50947;
XX
DT 18-MAY-1994 (first entry)
XX
DE Immunoglobulin binding protein gene derived from protein L gene.
XX
KW Peptide; immunoglobulin; binding; analysis; purification; ELISA;
KW enzyme linked immunoabsorbant assay; ss.
XX
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	repeat_unit	1..96
FT		/*tag= a
FT		/note= "Repeat units are not adjacent, repetitions of this sequence are not 100% homologous and begin at nucleotide positions 214, 427 and 649"
FT	repeat_unit	97..210
FT		/*tag= b
FT		/note= "Repeat units are not adjacent, repetitions of this sequence are not 100% homologous and begin at nucleotide positions 313, 535 and 757"

XX
PN WO9322439-A.
XX
PD 11-NOV-1993.
XX
PF 07-MAY-1993; 93WO-GB00950.
XX
PR 07-MAY-1992; 92GB-0009804.
PR 24-DEC-1992; 92GB-0026928.
XX
PA (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
XX
PI Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;
XX
DR WPI; 1993-368798/46.
DR P-PSDB; R42204.
XX
PT New immunoglobulin binding proteins derived from Protein L -
PT which bind immunoglobulin kappa light chains but not albumin or
PT cell walls
XX
PS Claim 13; Figure 2; 28pp; English.
XX
CC The synthetic immunoglobulin binding proteins derived from protein
CC L comprise repeated sequences from protein L which bind
CC immunoglobulin kappa light chains. They can be used in protein
CC analysis, purification procedures and other biochemical processes e.
CC g. ELISA. The synthetic molecules are of particular advantage if
CC they are free of regions in protein L which exhibit albumin and cell
CC wall binding.
XX
SQ Sequence 876 BP; 402 A; 141 C; 162 G; 171 T; 0 other;

Query Match 2.1%; Score 48.2; DB 14; Length 876;
Best Local Similarity 46.7%; Pred. No. 0.11;
Matches 189; Conservative 0; Mismatches 213; Indels 3; Gaps 1;

```
Qy   1061 atagtatcgcagagccaatcacttttaaggttgaagctggcaaagtgtatactattattg 1120
Db    362 aaaatggcgaatatacagcagacttagaagatggtggaaacacaatcaacattaaatttg 421

Qy   1121 atggaaaacagattgaaaatcccaataaagagatagtagagccttactcagtagaagcat 1180
Db    422 ctggaaaagaaacaccagaaacaccagaagaaccaaaagaagaagttacaatcaaagtta 481

Qy   1181 ataatgattttgaagaatttagcgttttaactacacaaaactatgcaaaattttattatg 1240
Db    482 acttaatctttgcagatggaaagatacaaacagcagaattcaaaggaacatttgaagaag 541

Qy   1241 caaaaaataaaaatggaagttcacaggttgtctattgctttaatgcagatctaaaatctc 1300
Db    542 caaca---gcaaaagcttatgcttatgcaaacttattagcaaaagaaaatggcgaatata 598

Qy   1301 caccagactctgaagatggtgggaaaacaatgactccagactttacaacaggagaagtaa 1360
Db    599 cagcagacttagaagatggtggaaacacaatcaacattaaatttgctggaaaagaaacac 658

Qy   1361 aatacactcatattgcaggtcgtgacctctttaaatatactgtgaaaccaagagataccg 1420
Db    659 cagaaacaccagaagaaccaaaagaagaagttacaatcaaagttaacttaatctttgcag 718

Qy   1421 atcctgacactttcttaaaacatatcaaaaaagtaattgagaaagg 1465
Db    719 atggaaaaacacaaacagcagaattcaaaggaacatttgaagaag 763
```

RESULT 10

```
Q50946
ID   Q50946 standard; cDNA; 3279 BP.
XX
AC   Q50946;
XX
DT   18-MAY-1994  (first entry)
XX
DE   Sequence encoding protein L.
XX
KW   Peptide; immunoglobulin; binding; analysis; purification; ELISA;
KW   enzyme linked immunoabsorbant assay; ss.
XX
OS   Peptococcus magnus.
XX
FH   Key             Location/Qualifiers
FT   CDS             103..3185
FT                   /*tag= a
FT                   /product= Protein L.
FT   repeat_unit     490..573
FT                   /*tag= b
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 673 and 856"
FT   repeat_unit     574..672
FT                   /*tag= c
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide position 757"
FT   repeat_unit     949..1044
FT                   /*tag= d
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 1162, 1375
FT                          and 1597"
FT   repeat_unit     1045..1158
FT                   /*tag= e
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 1261, 1483
FT                          and 1705"
FT   repeat_unit     1822..1938
FT                   /*tag= f
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 2347, 2545 and
FT                          2731"
FT   repeat_unit     2914..2934
FT                   /*tag= g
FT                   /note= "Repeat units are adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 2935, 2953,
FT                          2968, 2986, 3001, 3019 and 3034"
XX
PN   WO9322439-A.
XX
PD   11-NOV-1993.
XX
PF   07-MAY-1993;   93WO-GB00950.
XX
PR   07-MAY-1992;   92GB-0009804.
PR   24-DEC-1992;   92GB-0026928.
XX
PA   (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
XX
PI   Atkinson A,  Duggleby CJ,  Murphy JP,  Trowern AR;
XX
DR   WPI; 1993-368798/46.
DR   P-PSDB; R42203.
XX
PT   New immunoglobulin binding proteins derived from Protein L -
PT   which bind immunoglobulin kappa light chains but not albumin or
PT   cell walls
XX
PS   Disclosure; Figure 1; 28pp; English.
XX
CC   The synthetic immunoglobulin binding proteins derived from protein
CC   L comprise repeated sequences from protein L which bind
CC   immunoglobulin kappa light chains.  They can be used in protein
CC   analysis, purification procedures and other biochemical processes e.
CC   g. ELISA.  The synthetic molecules are of particular advantage if
CC   they are free of regions in protein L which exhibit albumin and cell
CC   wall binding (The repeat regions commencing at nucleotide numbers
CC   1045, 1261, 1483 and 1705).
XX
SQ   Sequence 3279 BP; 1505 A; 481 C; 625 G; 668 T; 0 other;


  Query Match             2.1%;  Score 48.2;  DB 14;  Length 3279;
  Best Local Similarity   46.7%;  Pred. No. 0.16;
  Matches  189;  Conservative    0;  Mismatches  213;  Indels    3;  Gaps    1;

Qy   1061 atagtatcgcagagccaatcacttttaaggttgaagctggcaaagtgtatactattattg 1120
           | | |  || | |  ||       ||  | | ||  |    || | |  | | ||   |||
Db   1310 aaaatggcgaatatacagcagacttagaagatggtggaaacacaatcaacattaaatttg 1369

Qy   1121 atggaaaacagattgaaaatcccaataaagagatagtagagccttactcagtagaagcat 1180
           |||||| | |    | |  |     ||||   |  |||        || |  |||
Db   1370 ctggaaaagaaacaccagaaacaccagaagaaccaaaagaagaagttacaatcaaagtta 1429

Qy   1181 ataatgattttgaagaatttagcgttttaactacacaaaactatgcaaaattttattatg 1240
           |       |||| |||    |   |   |||  || ||  | | | || |||| |  | |
Db   1430 acttaatctttgcagatggaaagatacaaacagcagaattcaaaggaacatttgaagaag 1489

Qy   1241 caaaaaataaaaatggaagttcacaggttgtctattgctttaatgcagatctaaaatctc 1300
           ||| |     ||| |    | |  | |    ||  |     || | | ||    ||| |
Db   1490 caaca---gcaaaagcttatgcttatgcaaacttattagcaaaagaaaatggcgaatata 1546

Qy   1301 caccagactctgaagatggtgggaaaacaatgactccagactttacaacaggagaagtaa 1360
           || ||||||  |||||||||||| || ||||| |      | ||| |   |  ||||  |
Db   1547 cagcagacttagaagatggtggaaacacaatcaacattaaatttgctggaaaagaaacac 1606

Qy   1361 aatacactcatattgcaggtcgtgacctctttaaatatactgtgaaaccaagagataccg 1420
           | | || |     | |      ||     ||| |   |  || ||   ||    | | |
Db   1607 cagaaacaccagaagaaccaaaagaagaagttacaatcaaagttaacttaatctttgcag 1666

Qy   1421 atcctgacactttcttaaaacatatcaaaaaagtaattgagaagg 1465
           ||    | ||      |  | |  |||||  |  | ||||  | |
Db   1667 atggaaaaacacaaacagcagaattcaaaggaacatttgaagaag 1711


RESULT  11
Q51556
ID   Q51556 standard; cDNA; 3279 BP.
XX
AC   Q51556;
XX
DT   18-MAY-1994  (first entry)
XX
DE   Sequence encoding protein L.
XX
KW   Protein; immunoglobulin; binding; immobilisation; light chains;
KW   antibodies; diagnosis; pharmaceutical; ss.
XX
OS   Peptococcus magnus.
XX
FH   Key             Location/Qualifiers
FT   CDS             103..3185
FT                   /*tag= a
FT                   /product= Protein L.
FT   repeat_unit     490..573
FT                   /*tag= b
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 673 and 856"
FT   repeat_unit     574..672
FT                   /*tag= c
```

```
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide position 757"
FT   repeat_unit     949..1044
FT                   /*tag=  d
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 1162, 1375
FT                          and 1597"
FT   repeat_unit     1045..1158
FT                   /*tag=  e
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 1261, 1483
FT                          and 1705"
FT   repeat_unit     1822..1938
FT                   /*tag=  f
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 2347 and 2545"
FT   repeat_unit     1939..2007
FT                   /*tag=  g
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 2479, 2665
FT                          and 2851"
FT   repeat_unit     2035..2094
FT                   /*tag=  h
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide position 2209"
FT   repeat_unit     2095..2208
FT                   /*tag=  i
FT                   /note= "Repeat units are not adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 2269"
FT   repeat_unit     2914..2934
FT                   /*tag=  j
FT                   /note= "Repeat units are adjacent, repetitions
FT                          of this sequence are not 100% homologous and
FT                          begin at nucleotide positions 2935, 2953,
FT                          2968, 2986, 3001, 3019 and 3034"
XX
PN   WO9322438-A.
XX
PD   11-NOV-1993.
XX
PF   07-MAY-1993;   93WO-GB00949.
XX
PR   07-MAY-1992;   92GB-0009804.
XX
PA   (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
XX
PI   Atkinson A,  Duggleby CJ,  Murphy JP,  Trowern AR;
XX
DR   WPI; 1993-368797/46.
DR   P-PSDB; R43699.
XX
PT   Immunoglobulin binding polypeptide, protein L - used for prodn.
PT   of pharmaceuticals and for immobilising antibodies e.g. on
PT   columns, in diagnostic tests and in assays
XX
PS   Disclosure; Figure 1; 29pp; English.
XX
CC   Protein L forms a complex with immunoglobulin Kappa light chain.
CC   Purified protein can be used as a reagent for immobilising
CC   antibodies e.g. on columns, in diagnostic tests and in assays.  It
CC   may also be used in the production of pharmaceuticals.
XX
SQ   Sequence 3279 BP; 1505 A; 480 C; 626 G; 668 T; 0 other;

  Query Match              2.1%;  Score 48.2;  DB 14;  Length 3279;
  Best Local Similarity   46.7%;  Pred. No. 0.16;
  Matches  189;  Conservative    0;  Mismatches  213;  Indels    3;  Gaps    1;

Qy    1061 atagtatcgcagagccaatcacttttaaggttgaagctggcaaagtgtatactattattg 1120
           | | |  || | |  ||       ||  | | ||  |    || | |  | | ||   |||
Db    1310 aaaatggcgaatatacagcagacttagaagatggtggaaacacaatcaacattaaatttg 1369

Qy    1121 atggaaaacagattgaaaatcccaataaagagatagtagagccttactcagtagaagcat 1180
           |||||||| | |    | |  |     ||||   |  |||        || |  |||
Db    1370 ctggaaaagaaacaccagaaacaccagaagaaccaaaagaagaagttacaatcaaagtta 1429

Qy    1181 ataatgattttgaagaatttagcgttttaactacacaaaactatgcaaaattttattatg 1240
           |       |||| |||    |   |   |||  || ||  | | | || ||||  |  | |
Db    1430 acttaatctttgcagatggaaagatacaaacagcagaattcaaaggaacatttgaagaag 1489

Qy    1241 caaaaaataaaaatggaagttcacaggttgtctattgctttaatgcagatctaaaatctc 1300
           ||| |     ||| |    | |  | |    ||  |     || | | ||    ||| |
Db    1490 caaca---gcaaaagcttatgcttatgcaaacttattagcaaaagaaaatggcgaatata 1546

Qy    1301 caccagactctgaagatggtgggaaaacaatgactccagactttacaacaggagaagtaa 1360
           || ||||||  |||||||||||| || ||||| |      | ||| |   |  ||||  |
Db    1547 cagcagacttagaagatggtggaaacacaatcaacattaaatttgctggaaaagaaacac 1606

Qy    1361 aatacactcatattgcaggtcgtgacctctttaaatatactgtgaaaccaagagataccg 1420
           | | || |     | |      ||     ||| |   |  || ||   ||    | | |
Db    1607 cagaaacaccagaagaaccaaaagaagaagttacaatcaaagttaacttaatctttgcag 1666

Qy    1421 atcctgacactttcttaaaacatatcaaaaaagtaattgagaagg 1465
           ||    | ||      |  | |  ||||  |  | ||||  | |
Db    1667 atggaaaaacacaaacagcagaattcaaaggaacatttgaagaag 1711


RESULT  12
C58017
ID   C58017 standard; DNA; 20674 BP.
XX
AC   C58017;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Arachidonic acid metabolism related genomic biallelic marker #651.
XX
KW   Human; biallelic marker; arachidonic acid metabolism; genotyping;
KW   detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;
KW   single nucleotide polymorphism; hybridisation assay; sequencing assay;
KW   specific amplification assay; identification; ERBM; 12-LO-RBM;
KW   eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.
XX
OS   Homo sapiens.
XX
PN   WO200047771-A2.
XX
PD   17-AUG-2000.
XX
PF   11-FEB-2000; 2000WO-IB00184.
XX
PR   12-FEB-1999;   99US-0119917.
PR   23-MAR-1999;   99US-0275267.
PR   07-MAY-1999;   99US-0133200.
XX
PA   (GEST ) GENSET.
XX
PI   Blumenfeld M,  Bougueleret L,  Chumakov I;
XX
DR   WPI; 2000-571881/53.
XX
PT   Novel biallelic markers useful for detecting conditions and genotypes
PT   associated with arachidonic acid metabolism -
XX
PS   Claim 67; Page 790-796; 802pp; English.
XX
CC   The present invention describes polynucleotides including biallelic
CC   markers derived from genes involved in arachidonic acid metabolism and
```

```
CC   from genomic regions flanking those genes. Methods from the present
CC   invention may be used to select individuals for clinical trials and
CC   predict responses to treatment with drugs. The polynucleotides may be
CC   used in hybridisation assays, sequencing assays and specific
CC   amplification assays for identifying an eicosanoid-related biallelic
CC   marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a
CC   segment of nucleotides containing an ERBM. The polynucleotides are
CC   useful in diagnostic kits. The markers may be used to detect conditions
CC   and genotypes associated with arachidonic acid metabolism. C57367 to
CC   C58018 and B24019 and B24020 represent sequences used in the
CC   exemplification of the present invention.
CC   N.B. Polymorphic bases (single nucleotide polymorphisms also known as
CC   SNPs) in the polynucleotide sequences from the present invention have
CC   been given as their corresponding degenerate bases e.g. a polymorphic
CC   base of C or T has been given as Y.
XX
SQ   Sequence 20674 BP; 5400 A; 5170 C; 4995 G; 5062 T; 47 other;


  Query Match             2.1%;  Score 48.2;  DB 21;  Length 20674;
  Best Local Similarity   45.9%;  Pred. No. 0.25;
  Matches  240;  Conservative    0;  Mismatches  278;  Indels    5;  Gaps    2;

Qy    1608 agacatgaatgatagtactttagcagttgctaaaatccttgtagaatacgctcaagatag 1667
           | |  | |||  |  || ||||  |  |  |||  |  |  || |||         |
Db   11084 atattttaatatttttaatttaataaatttttaatttaatattaaaattaaaatttttaatt 11143

Qy    1668 taatcctccacagctaactgaccttgatttctttattccgaataacaataaatatcaatc 1727
           ||||  |  |     |  | |  || |  |   || |   |||   ||||||| | |||
Db   11144 taatattaaattaaaatttttaatttttaaataaataattaaaatttaaataaatttaaatt 11203

Qy    1728 tcttattggaactcagtggcatccagaagatttagttgatattattcgtatggaagataa 1787
           |   |||  || | | |   |   | |   || | ||  || | ||  |||  ||   |
Db   11204 taaaatttaaatttaattttaaataaacatttaaatttttaattttaatattaaattaat 11263

Qy    1788 aaaagaagttatacctgtaactca---taatttaacattgagaaaaacggtgactggttt 1844
             ||     |||   | ||||| |   |||||||| ||| |   ||    | | |   ||
Db   11264 ttaatttaatattaatttaacttaatttaatttaatattaatttaa--cttaatttaatt 11321

Qy    1845 agctggtgacagaactaaagatttccattttgaaattgaattaaaaaataataagcaaga 1904
           |  |    |  || ||||  ||   ||||| || || |||   ||  ||||       |
Db   11322 aatttaatattaaattaaatttttaatatttttaattttaatattaatttaatttaattta 11381

Qy    1905 attgctttctcaaactgttaaaacagataaaacaaacctcgaatttaaagatggtaaagc 1964
           |||   |  | |    ||||     ||   |    | || |  |||||     |
Db   11382 atttaatattaattaatttaattttaatttttaattgcttcaatattaaattaaatttttaa 11441

Qy    1965 aaccattaatttaaaacatggggaaagtttaacacttcaaggtttaccagaaggttattc 2024
             |  ||| ||||| |       ||| ||| |   |  ||          ||  |||
Db   11442 tatttttatttaatattaaattaaaatttttaatttaaaattaaatatttaatattaaat 11501

Qy    2025 ttaccttgtcaaagaaacagattctgaaggctataaggttaaagttaatagccaagaagt 2084
           | |  |       |  |  | ||| | |  ||||        |  ||||||   || ||
Db   11502 taatttaaatttaatattaaattttaatagctaaggaaaatattttaatattaaaaaaag 11561

Qy    2085 agcaaatgctacagtttcaaaaacaggaataacaagtgatgag 2127
           | ||    ||    |    |    ||||| | || | || |||
Db   11562 accatggccttttttcaatagctaaggaaaagcaggagaagag 11604


RESULT  13
V74631
ID   V74631 standard; DNA; 5897 BP.
XX
AC   V74631;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #320.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    421..480
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    2221..2280
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    4021..4080
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
FT   misc_feature    5821..5880
FT                   /*tag= d
FT                   /note= "these bases represent a line of missing text in
FT                          the sequence listing in the specification. They
FT                          are included to maintain the nucleotide numbering
FT                          given in the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-0100117.
XX
PR   05-JAN-1996;   96US-0009861.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Fannon MR,  Kunsch CA;
PI   Rosen CA;
XX
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
XX
PS   Claim 1; Page 1178-1182; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
XX
SQ   Sequence 5897 BP; 2046 A; 852 C; 1114 G; 1642 T; 243 other;
```

```
  Query Match             2.1%;  Score 47.4;  DB 18;  Length 5897;
  Best Local Similarity   46.3%;  Pred. No. 0.27;
  Matches  156;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy     865 caaacgacttcagtacttattagaaagtatgctataggtgattactctaaattgcttgaa 924
           | ||  | | || ||  | ||  |    | |  | ||   || | | |  |    |
Db    4351 ccaattattgcattagattttgaatcaaaagaaaaagtaaatcaatttttagatttattt 4410

Qy     925 ggtgcaacattacagttgacaggggataacgtgaatagttttcaagcgagagtgtttagc 984
           | || | |||||   | | ||  | ||  |    |  || |||||   |
Db    4411 gatgaatcattattcgtaaaagtaggtatggaacttttttatcaagaaggtcctcaatta 4470

Qy     985 agtaatgatattggagaaagaattgaactatcagatggaacttatactttaactgaattg 1044
           | |||||| ||  |||||||    |      |  |  |  ||  |   |  |||  |
Db    4471 attaatgagataaaagaaagaggccatgatgtatttttagatttaaaactgcatgatatt 4530

Qy    1045 aattctccagctggttatagtatcgcagagccaatcacttttaaggttgaagctggcaaa 1104
             |  | ||| |||| |    || | ||  | |   |  || || |||||    |  ||
Db    4531 cctaatacagttggtaaggcgatggaaggactagctaaattgaatgttgatctggtaaat 4590

Qy    1105 gtgtatactattattgatggaaaacagattgaaaatcccaataaagagatagtagagcct 1164
           ||  || ||  |  ||  | ||||  |||     |  ||| |||||   ||  | | | |
Db    4591 gttcatgctgctggtggcgtaaaaatgatgtctgaggccattaaaggattaagaaaacat 4650

Qy    1165 tactcagtagaagcatataatgattttgaagaattta 1201
             |   ||    |  |  || ||   |   | |  |||
Db    4651 aatcaagatacaaaaattattgcagtaacacagctta 4687


RESULT  14
A70106
ID   A70106 standard; DNA; 7458 BP.
XX
AC   A70106;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:239.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 462-464; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC   protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 7458 BP; 3523 A; 514 C; 1174 G; 2247 T; 0 other;


  Query Match             2.1%;  Score 47.4;  DB 21;  Length 7458;
  Best Local Similarity   45.8%;  Pred. No. 0.29;
  Matches  201;  Conservative    0;  Mismatches  236;  Indels    2;  Gaps    1;

Qy    1544 tatattatttcactgatagtgctgaattagataaggataaactaaaagactatcatggtt 1603
           || | ||  | |  || |    | |||||  | ||  | ||   || |   | ||  ||
Db     657 taaaataaataaaggaaaaaaattaattaaaaaaaaaaaagtaaataatatttctaggag 716

Qy    1604 ttggagacatgaatgatagtactttagcagttgctaaaatccttgta--gaatacgctca 1661
             |||  |||  || |||  | | | | | |   | | ||  | | |  |   || |
Db     717 gaggaatcatatattatataaatgtcgtaataagttatataatggaaatgttttttctga 776

Qy    1662 agatagtaatcctccacagctaactgaccttgatttctttattccgaataacaataaata 1721
             |||| || |       |  | | ||    | |||  || | |     | | || || ||
Db     777 cgatattattaaaagtgaagtgaatgtatgtaattcgttaactgtacttcataagaatta 836

Qy    1722 tcaatctcttattggaactcagtggcatccagaagatttagttgatattattcgtatgga 1781
           | |   |  ||| |  | | | |   ||    |          ||| ||    |||  |
Db     837 taatattaatatggataattatttagatgataatatacatacaaataatagcaatattta 896

Qy    1782 agataaaaaagaagttatacctgtaactcataatttaacattgagaaaaacggtgactgg 1841
             ||||  ||  |    |          |  ||||| |||||   |  |       |  ||
Db     897 tgatattaattatacaaatgaaaatgttattaattcaacatgtcgttattatcccatagg 956

Qy    1842 tttagctggtgacagaactaaagatttccattttgaaattgaattaaaaaataataagca 1901
           |     |        || |||||||     |    ||       | |||||  ||||   |
Db     957 taataataacacattaagtaaagatgaagttacaaaatcaagttcaaaaattaattccct 1016

Qy    1902 agaattgctttctcaaactgttaaaacagataaaacaaacctcgaatttaaagatggtaa 1961
             |   ||  | | | | ||||   | ||||||   |  |     ||  | |||  |||
Db    1017 ttcttattttgatgatattattaatgtaaataaaaatgatatacctatttttacatgataa 1076

Qy    1962 agcaaccattaatttaaaa 1980
           || ||  || ||| ||| |
Db    1077 agaaaatataaatataata 1095


RESULT  15
V21209
ID   V21209 standard; DNA; 1664976 BP.
XX
AC   V21209;
XX
DT   10-NOV-1998  (first entry)
XX
DE   Methanococcus jannaschii circular chromosome.
XX
KW   Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW   genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS   Methanococcus jannaschii.
XX
PN   WO9807830-A2.
XX
```

PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-US14900.
XX
PR 22-AUG-1996; 96US-0024428.
XX
PA (GENO-) INST GENOMIC RES.
PA (UNII) UNIV ILLINOIS FOUND.
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
XX
DR WPI; 1998-169145/15.
XX
PT Complete genome sequence of methano-genic archaeon, Methanoccoccus
PT jannaschii - useful in identification of M. jannaschii genome
PT fragment
XX
PS Claim 13; Page 152-585; 614pp; English.
XX
CC The present sequence represents the complete 1.66-megabase pair genome
CC sequence of the Methanococcus jannaschii circular chromosome. The
CC present invention describes M. jannaschii open reading frames from the
CC genome sequence. The invention also describes a computer based system
CC for identifying fragments of the M. jannaschii genome that are
CC homologous to target nucleotide sequences, comprising: (a) data storage
CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
CC least 99.9% identical to it; (b) search means for comparing a target
CC sequence to the nucleotide sequence of the data storage means to
CC identify a homologous sequence, and (c) retrieval means for obtaining
CC the homologous sequence. The method, which is based on whole genome
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC of which consists of 3 physically distinct elements, a large circular
CC chromosome (the 1664976 bp sequence given in V21209), a large circular
CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a
CC small circular extra-chromosomal element (the 16550 bp sequence given in
CC V21211), can be used in the identification of M. jannaschii genome
CC fragment.
XX
SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Query Match 2.1%; Score 47.4; DB 19; Length 1664976;
Best Local Similarity 46.3%; Pred. No. 1.2;
Matches 156; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

```
Qy   1778 tggaagataaaaaagaagttatacctgtaactcataatttaacattgagaaaaacggtga 1837
          | ||  |  | || |||  | ||  | ||||| | || |||  |  |  ||||  | |
Db 201070 ttgataaatataaggaatatttaggtctaactgaaaaattagaagagcttaaaaatatta 201129

Qy   1838 ctggtttagctggtgacagaactaaagatttccattttgaaattgaattaaaaaataata 1897
            | |     ||  || |    |||    | | | | | | |||  |  || | |||| |
Db 201130 aagatgggcttgaagagatttataatatatgcaactctaagattttagcaatagataaca 201189

Qy   1898 agcaagaattgctttctcaaactgttaaaacagataaaacaaacctcgaatttaaagatg 1957
            |      |  |. | ||   | ||    |  | |   ||||  | || ||  ||| |
Db 201190 ttaagaggaagtataataaagaagatattgaaatttacctaaacaacaaaatttttagagg 201249

Qy   1958 gtaaagcaaccattaatttaaaacatggggaaagtttaacacttcaaggtttaccagaag 2017
          |||       ||||||   | | | |    ||  |      || |      ||  || |
Db 201250 ttaataaggaaattaatgatatagaagaaagaatatcctacattaaccaaaaacttgatg 201309

Qy   2018 gttattcttaccttgtcaaagaaacagattctgaaggctataaggttaaagttaatagcc 2077
                |||  ||   |||||    |      |    |      | ||  ||| || | |
Db 201310 aaataaactacaatgaagaagaacataaaaaaataaaagagctctatgaaaataagagac 201369

Qy   2078 aagaagtagcaaatgctacagtttcaaaaacaggaat 2114
          ||||| | |  || |  | ||    |||||||| |||
Db 201370 aagaacttgataacgtaagagaacaaaaaacagaaat 201406
```

Search completed: June 6, 2001, 21:48:17
Job time: 8176 sec